Pergamon

Insect Biochemistry and Molecular Biology 29 (1999) 1053–1063

Insect Biochemistry and Molecular Biology

www.elsevier.com/locate/ibmb

Characterization of midgut trypsin-like enzymes and three trypsinogen cDNAs from the lesser grain borer, *Rhyzopertha dominica* (Coleoptera: Bostrichidae)

Yu-Cheng Zhu, James E. Baker *

Grain Marketing and Production Research Center, ARS-USDA, 1515 College Avenue, Manhattan, KS 66502, USA

Received 10 March 1999; received in revised form 1 July 1999; accepted 20 July 1999

Abstract

Protein digestion in the lesser grain borer, *Rhyzopertha dominica* (F.) (Coleoptera: Bostrichidae), results from the action of a complex of serine proteinases present in the midgut. In this study we partially characterized trypsin-like enzyme activity against N-α-benzoyl-L-arginine *p*-nitroanilide (BA*p*NA) in midgut preparations and cloned and sequenced three cDNAs for trypsinogen-like proteins. BA*p*NAase activity in *R. dominica* midgut was significantly reduced by serine proteinase inhibitors and specific inhibitors of trypsin, whereas BA*p*NAase activity was not sensitive to specific inhibitors of chymotrypsin or aspartic proteinases. However, trans-epoxysuccinyl-L-leucylamido-(4-guanidino) butane (E-64) inhibited BA*p*NAase activity by about 30%. BA*p*NAase was most active in a broad pH range from about pH 7 to 9.5. The gut of *R. dominica* is a tubular tract approximately 2.5 mm in length. BA*p*NAase activity was primarily located in the midgut region with about 1.5-fold more BA*p*NAase activity in the anterior region compared to that in the posterior region. Proteinases with apparent molecular masses of 23–24 kDa that were visualized on casein zymograms following electrophoresis were inhibited by TLCK.

Three cDNAs for trypsinogen-like proteins were cloned and sequenced from mRNA of *R. dominica* midgut. The full cDNA sequences consisted of open reading frames encoding 249, 293, and 255 amino acid residues for *Rdo*T1, *Rdo*T2, and *Rdo*T3, respectively. cDNAs *Rdo*T1, *Rdo*T2, and *Rdo*T3 shared 77–81% sequence identity. The three encoded trypsinogens shared 54–62% identity in their amino acid sequences and had 16–18 residues of signal peptides and 12–15 residues of activation peptides. The three predicted mature trypsin-like enzymes had molecular masses of 23.1, 28, and 23.8 kDa for *Rdo*T1, *Rdo*T2, and *Rdo*T3, respectively. Typical features of these trypsin-like enzymes included the conserved N-terminal residues $IVGG^{62–65}$, the catalytic amino acid triad of serine proteinase active sites (His^{109}, Asp^{156}, Ser^{257}), three pairs of conserved cysteine residues for disulfide bridges, and the three residues (Asp^{251}, Gly^{274}, Gly^{284}) that determine specificity in trypsin-like enzymes. In addition, *Rdo*T2 has both a PEST-like sequence at the C-terminus and a free Cys^{158} near the active site, suggesting instability of this enzyme and/or sensitivity to thiol reagents. The sequences have been deposited in GenBank database (accession numbers AF130840 for *Rdo*T1, AF130841 for *Rdo*T2, and AF130842 for *Rdo*T3). Published by Elsevier Science Ltd.

Keywords: Lesser grain borer; *Rhyzopertha dominica*; Digestion; Serine proteinase; Trypsin; BA*p*NAase; cDNA; Cloning; Sequence

1. Introduction

Application of chemical protectants to grains and grain products for insect pest management during storage is being supplanted by more environmentally-sensitive technologies. Several of these replacement technologies include improved sampling protocols for monitoring pest insect populations and the increased use of aeration and fumigation when appropriate (Hagstrum and Subramanyam, 1996). Another technology that can help maintain grain quality during storage is the development of transformed cereals containing bioactive proteins that suppress insect growth and reduce pest insect population size (Baker and Kramer, 1996) and that can subsequently reduce the need for chemical applications. Such resistant cereals offer a passive technology that could be easily integrated with all control strategies currently available in the storage ecosystem.

* Corresponding author. Tel.: +1-785-776-2785; fax: +1-785-776-2792.

E-mail address: baker@usgmrl.ksu.edu (J.E. Baker)

0965-1748/99/$ - see front matter. Published by Elsevier Science Ltd.
PII: S0965-1748(99)00081-8

Bioactive proteins, such as digestive proteinase inhibitors, engineered into wheat would be targeted against the lesser grain borer, *Rhyzopertha dominica* (F.), and the *Sitophilus* spp. grain weevils, the most economically-important coleopterans that attack intact kernels. To design and incorporate an inhibitor complex that would adversely affect both species, knowledge of the biochemistry of protein digestion in both species, including relevant differences between the species, is necessary. Such information is even more critical in view of the known complexities of enzyme:inhibitor interactions within the insect gut. These complexities include the potential for adaptation to the presence of dietary inhibitors by production of endoproteinases insensitive to a given inhibitor (Broadway, 1995; Jongsma et al., 1995) and the presence of multi-gene families of proteinases with differential sensitivity of enzymes within the same mechanistic class to a given inhibitor (Bown et al., 1997).

More information is available on digestion in *Sitophilus* spp. compared with that available for *R. dominica*. Although serine proteinases are present in low levels in the three major *Sitophilus* spp (Baker, 1982), cysteine proteinases are the predominant mechanistic class (Murdock et al., 1987; Liang et al., 1991; Houseman and Thie, 1993). In *S. zeamais* Motschulsky, a gene family encoding at least 4 cathepsin L-like cysteine proteinases with different properties has been characterized (Matsumoto et al. 1997, 1998).

Protein digestion in *R. dominica* is less-well studied but evidence indicates that, in contrast to the activity of cysteine proteinases in *Sitophilus* spp., serine proteinases, including significant hydrolytic activity against α-benzoyl-DL-arginine-*p*-nitroanilide (BA*p*NA), are predominant in this bostrichid beetle (Konarev and Fomicheva, 1991; Oppert et al., unpublished). The present paper reports additional studies on the proteinases in the *R. dominica* midgut including partial characterization of BA*p*NA-hydrolyzing enzymes and the cloning and sequencing of three trypsinogen-like cDNAs from this major pest of stored grain.

2. Materials and methods

2.1. Insect cultures

Stock culture of *R. dominica* were maintained on hard red winter wheat at 27°C and 55–65% RH. Adults that had emerged for 1–3 wk were generally used in these studies.

2.2. Preparation of midgut homogenates

Intestinal tracts (about 2.5 mm in length) were dissected in cold saline (128 mM NaCl, 4.7 mM KCl, 2.8 mM $CaCl_2$) by holding and pressing the abdomen and by removing the head simultaneously with forceps. Midgut enzyme solution was prepared as described by Oppert et al. (1994). After dissection, midguts were immersed in ice-cold 1 mM dithiothreitol (DTT) at a ratio of 1 gut/5 μl buffer. The mixture was vortexed briefly and centrifuged for 2 min at 15,000 ***g***. Supernatant was transferred to a new tube and frozen at −20°C until assayed.

2.3. Proteinase inhibitors

Phenylmethylsulfonyl fluoride (PMSF), pepstatin, *trans*-epoxysuccinyl-L-leucylamido-(4-guanidino) butane (E-64), benzamidine, potato tuber carboxypeptidase inhibitor (PCPI), TPCK, N-α-tosyl-L-lysine chloromethyl ketone (TLCK), and soybean trypsin inhibitor (STI), were obtained from Sigma Chemical Company. Stock solutions were prepared in ethanol for pepstatin and TPCK, and in H_2O for the remaining inhibitors. Inhibitors (1 μl) were preincubated with 5 μl enzyme solution and 45 μl 100 mM tris-chloride pH 8 for 15 min at 30°C. Residual BA*p*NAase activity was determined by addition of 50 μl BA*p*NA in 100 mM tris-chloride pH 8 (final BA*p*NA concentration was 0.5 mg/ml). Final inhibitor concentrations are given in the Results. Mean residual activity against BA*p*NA was based on assays with two groups of ~150 midguts with triplicate analyses for each inhibitor with each extract. A molar extinction coefficient of 8800 $mole^{-1}$ l^{-1} for *p*-nitroaniline was used to determine specific activity.

2.4. pH optimum

The universal buffer system (pH 2–11.5) of Frugoni (1957) was used to determine optimum pH for BA*p*-NAase activity. In the assay, 5 μl enzyme solution (equivalent to enzyme from one midgut) was mixed with 50 μl universal buffer at each pH. The reaction was started by adding 45 μl of BA*p*NA in 1% aqueous N,N-dimethylformamide (final concentration of BA*p*NA was 0.5 mg/ml). Absorbance at 405 nm was monitored for 10 min at 37°C. Blanks were included at each pH to correct for endogenous hydrolysis of substrate. Results are based on two groups of midguts with triplicate analyses at each pH in the range pH 2 to pH 11.5.

2.5. Localization of proteinase activity

Intact intestinal tracts were divided into four regions: foregut, anterior midgut, posterior midgut, and hindgut and homogenates prepared as above. Enzyme solution (5 μl) for each individual gut section was mixed with 95 μl of BA*p*NA solution in 100 mM tris-chloride pH 8 (final BA*p*NA concentration was 0.5 mg/ml) in a microtiter plate. Absorbance at 405 nm was monitored.

Mean activity was based on results with gut sections dissected from five midguts.

2.6. Zymogram analysis

Five μl midgut enzyme solution was mixed with 1 μl of 100 mM TLCK in distilled H_2O and incubated at 30°C for 2 hr in a capped microcentrifuge tube. The preincubation concentration of TLCK was 17 mM. After the preincubation, each sample was mixed with an equal volume of 2× sample buffer (0.125 mM tris-chloride pH 6.8, 20% glycerol, 4% SDS, 0.005% bromophenol blue) and loaded onto a 4–16% tris-glycine gel containing blue-stained β-casein (Novex, San Diego, CA). Following electrophoresis, enzymes were renatured in 2.5% Triton X-100, and the gel was developed by following manufacturer's instructions.

2.7. Cloning trypsin-like protein cDNA

Approximately 0.5 g of adult *R. dominica* was ground in liquid nitrogen. Total RNA was extracted with guanidine thiocyanate solution and precipitated with isopropanol (Titus, 1991). The poly(A) RNA was isolated from the total RNA using PolyATtract mRNA isolation system (Promega, Madison, WI). To obtain the 3′-end of trypsin-like cDNA, reverse transcription was conducted, and amplification of 3′-end of cDNA was performed using the 3′-RACE (rapid amplification of cDNA end) system (Gibco BRL Life–Technologies, Gaithersburg, MD).

Polymerase chain reaction (PCR) was carried out with an oligo-dT reverse primer and a forward degenerate primer, 5′-TGYCARGGNGAYWSNGGNGGNCCNYT-3′, designed from a highly conserved region (CQGDSGGPL) in both *Manduca sexta* trypsin and chymotrypsin cDNAs located approximately 250 bp from the 3′ end (Peterson et al. 1994, 1995). PCR-amplified DNA fragments (~250 bp) were cloned into a p-GEM-T vector (Promega, Madison, WI). Sequences of these clones were determined by using an automated sequencer. The cDNA sequence of the trypsin-like proteins was confirmed by homology search of GenBank provided by the National Center for Biotechnology Information using Blastx protocol (Altschul et al., 1990; Gish and States, 1993).

Three clones carried trypsin-like cDNA fragments with different sequences. To obtain the full sequences of these three trypsin-like cDNAs, two reverse primers were designed for each clone based on the corresponding 3′-end sequence of each cDNA (Fig. 2). Reverse transcription for each trypsin-like clone was performed by using the reverse primer and mRNA isolated from midgut tissue dissected from 300 adults. One semi-nested amplification was performed by using the 5′-RACE system (Gibco BRL Life–Technologies) with a forward abridged anchor primer and an appropriate reverse primer for each clone. DNA fragments (′650 bp) resolved from 5′-RACE amplification were cloned into a p-GEM-T vector and sequenced from both directions by using vector primers and appropriate primers designed from the obtained cDNA sequence.

To verify the cDNA sequence of trypsinogen that was determined by using Taq DNA polymerase, a thermostable proof-reading Pfu DNA polymerase (Promega) was used to reamplify a full length trypsinogen cDNA fragment from 3′-RACE cDNA by using a forward primer designed from the cDNA sequence determined with the Taq polymerase and an oligo-dT reverse primer. This fragment was A-tailed and cloned into a pGEM-T vector. The sequence of the insert was determined from both directions.

The Wisconsin Sequence Analysis Package GCG Unix version 9.0 (Genetics Computer Group, Madison, WI) including Pileup, Gap, Motifs, Distances, and Growtree programs was used to analyze the similarity of trypsinogen sequences (gap weight=3, gap length weight=1). Sequence analysis tools of the ExPASy Molecular Biology Server of Swiss Institute of Bioinformatics were used to process data of deduced protein sequences.

3. Results

3.1. Sensitivity of proteinases to inhibitors

Inhibitors of serine proteinases, or of trypsin, including aprotinin, leupeptin, STI, TLCK, PMSF, and benzamidine inhibited more than 88% of BA*p*NAase activity in *R. dominica* midguts (Table 1). Chymostatin, a chymotrypsin inhibitor, and E-64, thought to be specific for cysteine proteinases but also able to inhibit trypsin-like activity depending on the substrate used (Sreedharan et al., 1996), were moderately inhibitory (30–34%). Inhibition of less than 10% was obtained with TPCK, cystatin, PCPI and pepstatin. The effect of these latter four inhibitors was not significantly different (P=0.05) from uninhibited activity.

3.2. Influence of pH on BApNAase activity

BA*p*NAase activity in midguts of adult *R. dominica* increased gradually to a broad maximum level between pH 6.5 and 9.5 and then declined (Fig. 1A). BA*p*NAase activities were less than one third of maximal activity below pH 4.5.

3.3. Distribution of proteinase activity

Of total BA*p*NAase activity recovered, 54% was located in the anterior region of the midgut, and 30% in

Table 1
Effect of selected proteinase inhibitors on BA*p*NAase activity in extracts from midguts of *R. dominica*

Inhibitors	Conc.	Target Enzyme	Hydrolysis of BA*p*NA pmol/min/gut	Relative activity[a]	
Control	–	–	51.8	100	ab
Aprotinin	10 μM	Serine protease	1.4	3	d
PMSF	5 mM	Serine protease	4.2	8	d
STI	100 μM	Chymo-/trypsin	2.0	4	d
Chymostatin	10 μM	Chymotrypsin	36.1	70	c
TPCK	1 mM	Chymotrypsin	46.8	90	b
TLCK	1 mM	Trypsin	2.9	6	d
Leupeptin	200 μM	Serine/cysteine	1.7	3	d
E-64	100 μM	Cysteine	34.1	66	c
Cystatin	2 μM	Cysteine	48.5	94	b
Pepstatin	2 μM	Aspartic protease	56.6	109	a
Benzamidine	10 mM	Arg. bind pocket	6.4	12	d
PCPI	5 μM	Carboxypeptidase	47.8	92	b

[a] Means followed by same letters are not significantly different (P=0.05).

the posterior region of the midgut. The remaining activity (ca 16%) was in the hindgut. This latter activity may be from contaminating midgut fluid passing into the hindgut during dissection. There was no measurable BA*p*NAase activity in the foregut, which includes a small crop region.

3.4. Proteinase zymograms

Seven bands with proteinase activity (P1–P7) from midgut extracts of *R. dominica* were resolved in this gel system (Fig. 1B, lane 2). When midgut extracts were preincubated with TLCK and electrophoresed, proteinase P7 with a size of ~23 kDa and the minor proteinases P1–P5 that are larger than 30 kDa (lane 1) were completely suppressed by TLCK. Complete inhibition of P7 indicates that the activity associated with this band may be trypsin-like. Activity in P6 was partially inhibited by TLCK.

3.5. Trypsinogen-like protein cDNA

In the 3′-RACE amplifications, a degenerate primer was used, which was matched to cDNA sequences corresponding to the active site of both trypsin and chymotrypsin-like enzymes. A total of 25 recombinant clones from 3′-RACE was sequenced. Four clones carried inserts with deduced protein sequences that matched trypsin-like protein sequences in the GenBank. Two of the trypsin-like cDNAs contained identical sequences. Full cDNA sequences for the three different trypsinogen-like clones were successfully amplified and sequenced from midgut mRNA using RT–PCR and 5′-RACE cloning protocols.

The three trypsinogen-like cDNAs *Rdo*T1, *Rdo*T2, and *Rdo*T3, amplified by using Pfu DNA polymerase, contained 816, 940, and 850 nucleotides, respectively (Figs. 2A to C). Each of these cDNAs contained a long open-reading frame, the polyadenylation signal AATAAA, and a poly A tail. The open reading frames consisted of 747, 879, and 765 nucleotides for clones *Rdo*T1, *Rdo*T2, and *Rdo*T3, respectively.

3.6. Deduced protein sequences

The trypsinogen-like proteins encoded by clones *Rdo*T1, *Rdo*T2, and *Rdo*T3, are 249, 293, and 255 amino acid residues long, respectively, and the corresponding signal/activation peptides predicted by using Signal 1P software (Nielsen et al., 1997) contained 18/12, 17/15, and 16/13 amino acid residues. The calculated molecular masses for the predicted mature enzymes were 23.1, 28, and 23.8 kDa for clones *Rdo*T1, *Rdo*T2, and *Rdo*T3, respectively, and the calculated p*I*s corresponding to each of the predicted mature enzymes were 4.37, 3.85, and 4.1, respectively. Based on multiple sequence alignment (Fig. 3), the C terminus of *Rdo*T2 differed from those of the other two enzymes in having a very high percentage of proline (34%) and glutamic acid (31%) residues.

The three predicted amino acid sequences encoded by the *R. dominica* trypsinogen-like cDNAs were aligned with seven highly homologous insect trypsinogen-like proteinases from GenBank (Table 2, Fig. 3). These trypsinogens contained the conserved $IleValGlyGly^{62–65}$ sequence that is preceded by an arginine or lysine and that is located at the N termini of most trypsinogens (numbers are assigned based on total number of 337 residues of multiple alignment of 10 insect trypsinogens). In this alignment, the three putative active site residues, His^{109}, Asp^{156}, and Ser^{257} were conserved in all ten insect trypsinogens. Six cysteine residues that were conserved in all of these insect trypsinogen-like proteins were located at positions 94, 110, 225, 241, 253, and 277.

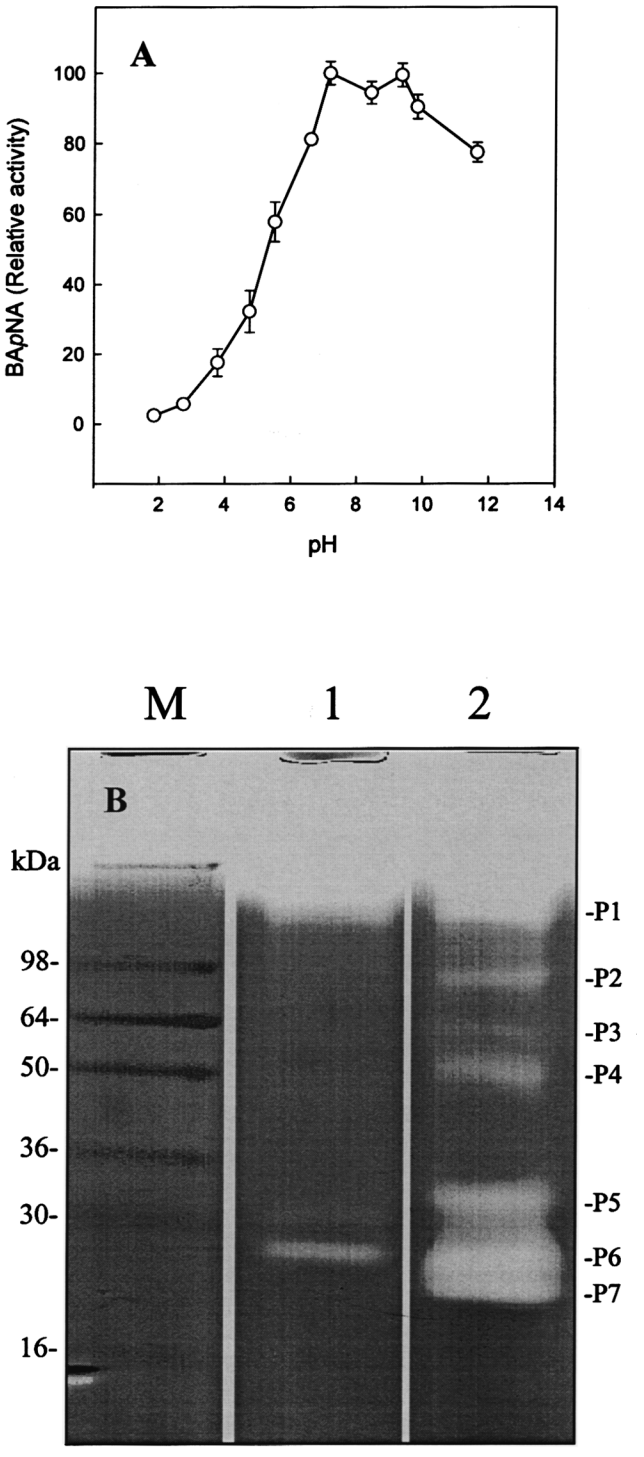

Fig. 1. (A) Effect of pH on BA*p*NAase activity (pmol/min/gut) of midgut preparations of *R. dominica* (values are means±SE and expressed relative to maximal activity); (B) Effect of preincubation of *R. dominica* midgut extract with TLCK prior to electrophoresis on casein-containing gels. Lane M: molecular weight marker proteins; lane 1, midgut enzymes preincubated with TLCK; lane 2, midgut enzymes without inhibitor. P1 to P7 designated as proteinases 1 to 7.

These cysteine residues are predicted to occur in disulfide bridge configurations (Wang et al., 1993). In addition to the conserved cysteine residues, the deduced amino acid sequence of *Rdo*T2 contained a seventh cysteine residue at position 158 (Fig. 3).

4. Discussion

4.1. Biochemical aspects of digestion in R. dominica

Characterization of the mechanistic classes of proteinases is complicated by a lack of distinct specificity of inhibitors as well as the differential susceptibility or specificity exhibited by different trypsin and chymotrypsin substrates (Johnston et al., 1995). However, activity measurements against BA*p*NA in *R. dominica*, and inhibition of activity against this specific substrate by selected serine proteinase inhibitors provides evidence supporting Konarev and Fomicheva (1991) and Oppert et al. (unpublished) that serine proteinases are actively involved in protein digestion in this stored product beetle. We have no evidence that aspartic or cysteine proteinases contribute significantly to the digestive process (Zhu and Baker, unpublished).

R. dominica is phylogenetically placed among wood eating species in the superfamily Bostrichoidea, distinct and in another evolutionary branch from the seed-eating *Sitophilus* spp. weevils in the Cuculionoidea. Although the two species feed and are major pests in an identical ecosystem (i.e. most stored cereals), digestion of dietary protein involves a complex of qualitatively different proteinases in the two species, serine proteinases in *R. dominica* and cysteine proteinases in *Sitophilus* (see Introduction). These data support the hypotheses of Terra (1988) and Terra and Ferreira (1994) concerning the importance of phylogeny rather than adaptation to diet itself in the evolution of digestive processes.

Hydrolases responsible for initial, intermediate, and final digestion of dietary protein in insects are morphologically arranged within and along the midgut (Terra and Ferreira, 1994). We assume the endoproteinase activity in *R. dominica* is from the luminal fluid or endoperitrophic region of the adult midgut. However, because of the small size of the gut in this species, we could only divide the midgut into anterior and posterior regions. In larvae of *Manduca sexta*, chymotrypsin mRNA was expressed more in the anterior and middle portions of the midgut, whereas trypsin-like mRNA was higher in the middle and posterior sections (Peterson et al., 1995). Because there are at least three trypsin genes present in *R. dominica*, additional studies will be required to determine if different genes are expressed differentially in different regions of the midgut. In addition, the anterior region of the midgut of *R. dominica* is slightly more acidic than the posterior region, ranging from about pH

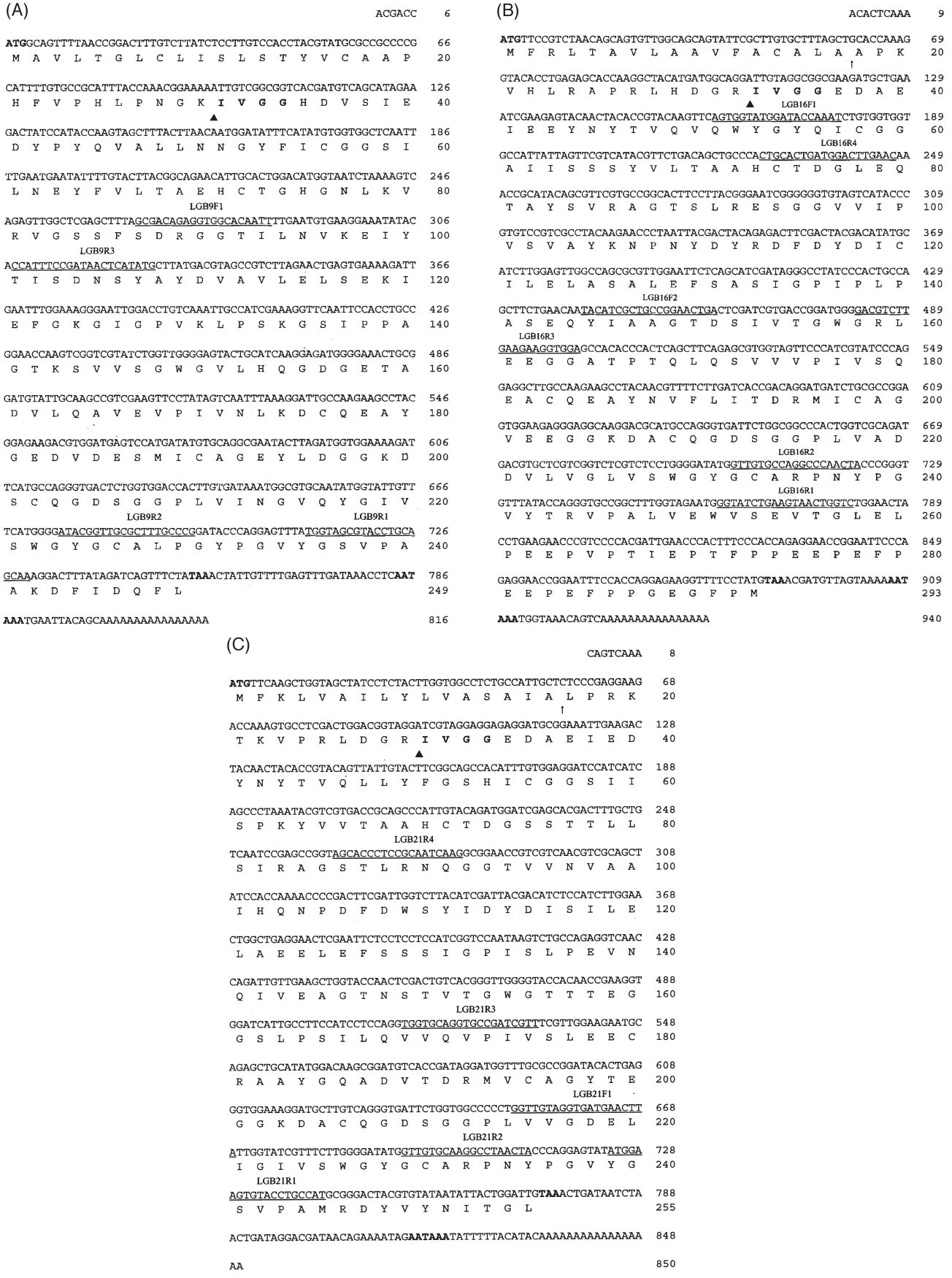

Fig. 2. Nucleotide and deduced amino acid sequences of trypsinogen-like cDNA isolated from *R. dominica* midgut RNA. (A) cDNA sequence of *Rdo*T1; (B) cDNA sequence of *Rdo*T2; (C) cDNA sequence of *Rdo*T3. ATG=start codon; TAA=termination codon; AATAAA=polyadenylation signal; ↑=predicted signal peptide cleavage site; ▲=predicted activation peptide cleavage site. IVGG are conserved N-terminal residues. Primer sequences used for cloning and sequencing cDNA are underlined and labeled on the top.

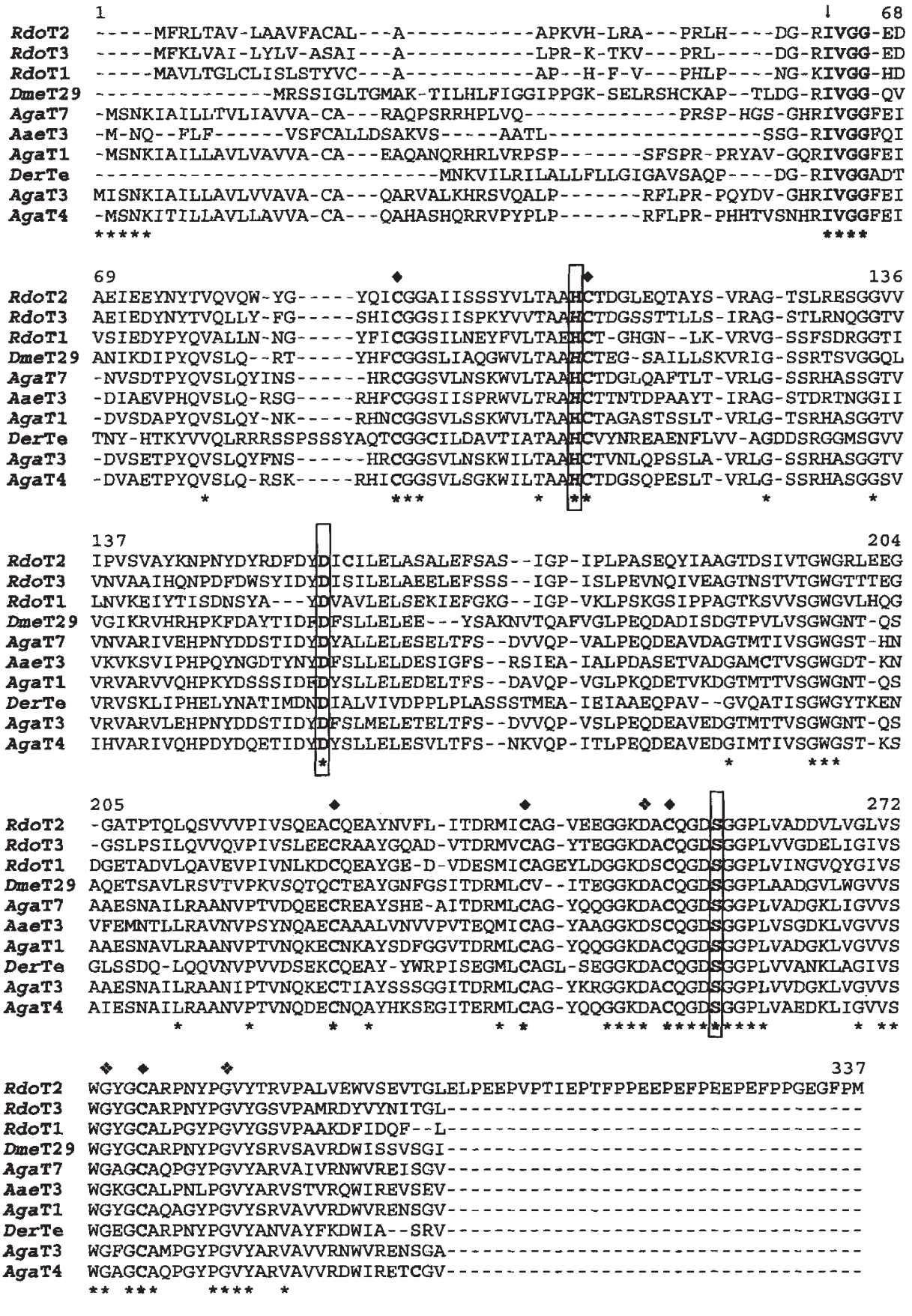
RdoT2
RdoT3
RdoT1
DmeT29
AgaT7
AaeT3
AgaT1
DerTe
AgaT3
AgaT4

Fig. 3. Predicted amino acid sequences of trypsinogen-like proteins from *R. dominica* and alignment with seven other insect trypsinogen-like sequences. *Rdo*T1=*R. dominica* trypsin precursor 1; *Rdo*T2=*R. dominica* trypsin precursor 2; *Rdo*T3=*R. dominica* trypsin precursor 3; *Dme*T29=trypsin-like proteinase Try29F of the fruit fly *D. melanogaster; Aga*T7=trypsin 7 precursor of African malaria mosquito *A. gambiae; Aae*T3=trypsin 3A1 precursor of yellow fever mosquito *A. aegypti; Aga*T1=trypsin 1 precursor of African malaria mosquito *A. gambiae; Der*Te=trypsin ETA precursor of fruit fly *D. erecta; Aga*T3-=trypsin 3 precursor of African malaria mosquito *A. gambiae*; *Aga*T4=-trypsin 4 precursor of African malaria mosquito *A. gambiae.* Functionally important residues are boxed and indicated with bold letters. Cysteines corresponding to the sites of the predicted disulfide bridges are marked with bold letters and solid diamonds (◆) on the top. Trypsin specificity determinant residues are indicated with (❖) on the top of sequences. Identical residues among all ten sequences are indicated with stars (★) at the bottom of sequences. The arrow (↓) indicates the N-terminal residues of the active enzymes. Hyphen represents sequence alignment gaps.

4.2. Molecular aspects of digestive proteinases

In this study we cloned 3 cDNAs from *R. dominica* midgut that encode trypsinogen-like proteins. All three cDNAs encode a sequence IleValGlyGly (Figs. 2 and 3) which is highly conserved in many trypsin- and chymotrypsin-like proteinases and marks the N-termini of the active enzymes (Jany and Haug, 1983; Wang et al., 1993). The cDNA *Rdo*T2 encodes a 35 residue C terminus which is not conserved with other insect trypsinogens (Fig. 3). The C terminal region is rich with proline (34%) and glutamic acid (31%) residues, which is similar to the PEST sequences thought to regulate a rapid degradation of the protein (Maekawa et al., 1998; Lee et al., 1997; Langenfeld et al., 1997) or to mediate phosphorylation of the protein (Marchal et al., 1998). It is possible that an instability of *Rdo*T2 because of the PEST-like sequence associated with this enzyme may be the reason that there was no distinct hydrolysis of casein on the zymograms at the 28 kDa region (Fig. 1B).

Table 2
Phylogenic relationship (GCG) between trypsinogen-like proteins of *R. dominica* and seven highly similar trypsinogen-like proteins from other insect species

Species	Trypsin precursor	Identity to *Rdo*T1 (%)	Identity to *Rdo*T2 (%)	Identity to *Rdo*T3 (%)	Phylogenic tree
R. dominica	*Rdo*T2	54	–	62	
R. dominica	*Rdo*T3	56	62	–	
R. dominica	*Rdo*T1	–	54	56	
A. aegypti	3A1	50	49	49	
D. erecta	ETA	49	50	48	
D. melanogaster	Try29F	50	55	53	
A. gambiae	1	47	52	55	
A. gambiae	3	46	51	54	
A. gambiae	7	49	57	54	
A. gambiae	4	46	52	52	

5.2 to 7.2 (Sinha, 1959), but it is not known how endogenous proteinase activity might be affected by these differences.

A total of seven activity bands were detected with a casein zymogram following electrophoresis of midgut preparations from *R. dominica*. Four minor bands (P1–P4, Fig. 1B) with molecular masses above 50 kDa and three major bands (P5–P7) with molecular masses estimated to be between 23 and 32 kDa were resolved in the gel system. P5 and P6 are partially to completely suppressed by preincubation of sample with TPCK (Zhu and Baker, unpublished). In the current study, P5 and P7 were totally suppressed and P6 was partially suppressed by TLCK, suggesting that trypsin-like proteinases contributed most activity to P7 and that both trypsin and chymotrypsin-like proteinases may contribute to P6. P6 and P7 had similar migration rates in this gel system which corresponded to molecular masses estimated to be about 23 kDa. The size of these proteinases matched molecular masses of chymotrypsin(23.8 kDa) (Zhu and Baker, unpublished) and *Rdo*T1 (23.1 kDa) and *Rdo*T3 (23.9 kDa) deduced from cDNAs cloned in this study.

A scan of the PROSITE database revealed that the deduced protein sequences of these cDNAs from *R. dominica* belong to the trypsin family of serine proteases. The same results were obtained with a search of the GenBank database by using the Blastx non-redundant program. From this latter search, similar sequences included trypsin-like proteinase Try29F of the fruit fly *Drosophila melanogaster* (Paululat, 1996), trypsin 7 precursor of African malaria mosquito *Anopheles gambiae* (Müller et al., 1993), trypsin 3A1 precursor of yellow fever mosquito *Aedes aegypti* (Kalhok et al., 1993), trypsin 1 precursor of African malaria mosquito *A. gambiae* (Müller et al., 1993), trypsin ETA precursor of fruit fly *D. erecta* (Wang and Hickey, 1996), trypsin 3 precursor of African malaria mosquito *A. gambiae* (Müller et al., 1993), and trypsin 4 precursor of African malaria mosquito *A. gambiae* (Müller et al., 1993).

GCG Gap analysis indicated that the deduced trypsinogen-like protein sequence *Rdo*T2 from *R. dominica* was most similar to the *Rdo*T3 with 71% sequence simi-

larity and 62% identity, and to the trypsin 7 precursor of African malaria mosquito *A. gambiae* with 66% similarity and 57% identity in the amino acid residues (Table 2). Sequence analyses using GCG Distances and Growtree methods also indicated that the trypsinogen-like protein sequence from *R. dominica* was very similar to the trypsin 3A1 precursor of yellow fever mosquito *A. aegypti* and the trypsin ETA precursor of the fruit fly *D. erecta* (Table 2).

The predicted amino acid sequences encoded by the three *R. dominica* cDNAs were aligned with seven homologous insect trypsin-like proteases (Fig. 3). These sequences contained all of the conserved residues representing typical features for trypsin proteinases. The putative active site residues, His^{109}, Asp^{156}, Ser^{257} (Fig. 3), which form the catalytic triad in serine proteases are conserved in all of these sequences (Kraut, 1977; Wang et al., 1993; Peterson et al., 1994). Six cysteine residues, predicted to occur in disulfide bridge configurations among trypsins and chymotrypsins, exhibited the same conserved pattern of cysteine residues found in all ten insect serine proteinases. The residues (Asp^{251}, Gly^{274}, Gly^{284}) which define the substrate binding pocket also were highly conserved in all of these trypsin-like enzymes. In these enzymes, Asp^{251} is predicted to be located at the bottom of the binding pocket. The Asp residue at this position determines specificity in both invertebrate and vertebrate trypsins by stabilizing the Lys or Arg residue at the substrate cleavage site through ionic interactions (Hedstrom et al., 1992; Wang et al., 1993).

Among the three deduced trypsin protein sequences of *R. dominica*, *Rdo*T1, *Rdo*T2 and *Rdo*T3 had 6, 7, and 6 cysteine residues, respectively. Besides three pairs of conserved cysteine residues, a Cys^{158} was present in *Rdo*T2, which was located near the catalytic triad, a putative novel position for serine proteases (Sasaki et al., 1993).

It is very likely that Cys^{158} in *Rdo*T2 is a free cysteine residue. Jongsma et al. (1996) and Gatehouse et al. (1997) suggested that a free cysteine in the active site or elsewhere on the protein made it sensitive to thiol-reacting inhibitors. In the current study, approximately 34% of BA*p*NAase activity was suppressed by E-64, whereas only 6% of the activity was inhibited by cystatin. The presence of a free Cys^{158} in *Rdo*T2 of *R. dominica* may make this enzyme sensitive to E-64, which has recently been shown to also inhibit trypsin activity against BA*p*NA (Sreedharan et al., 1996).

Trypsin gene families have been found in other insects including three cDNAs encoding alkaline midgut trypsins cloned from *M. sexta* (Peterson et al., 1994), and seven trypsinogen cDNAs identified in *A. gambiae* (Müller et al., 1993). In mosquitos, an *early* trypsin may play a role in initiating a cascade of events leading to the subsequent expression of *late* trypsins (Barillas-Mury et al., 1995), and different trypsins may represent the functional homologues of the *early* trypsin (Müller et al., 1995). Wang et al. (1995) suggested that ancestors of dipterans and lepidopterans may have had only one trypsin gene, but that some species may have gained extra copies by subsequent gene duplications.

The adaptive value of multigene trypsin families is not clear. Multigene trypsin families may have evolved to provide a more efficient mechanism for protein digestion as well as to provide an adaptive advantage for phytophagous species feeding on plants that contain proteinase inhibitors (Bown et al., 1997; Reeck et al., 1998). Multigene families may allow the induction of proteinases that are insensitive to dietary inhibition, as noted by Jongsma et al. (1995), Broadway (1995) and Bolter and Jongsma (1995), or initiate proteolysis of proteinase inhibitors by non-target digestive proteinases (Michaud et al., 1995). We have confirmed that *R. dominica* has evolved a complex of at least three trypsin-like genes involved in protein digestion. Because *S. zeamais* has a corresponding family of at least four cysteine proteinases that may be involved in digestion (Matsumoto et al., 1997), successful development of transformed cereal cultivars resistant to both pest species will require not only the presence of stacked cysteine and serine proteinase inhibitors, but knowledge of the sensitivity of individual members within each proteinase family in both insect species to each inhibitor selected.

Acknowledgements

The authors wish to thank Haobo Jiang of Kansas State University, and B. Oppert and K. Kramer of USDA–ARS–GMPRC Manhattan for their technical advice. The authors are grateful to K. Kramer, S. Muthukrishnan, and K.Y. Zhu for reviewing an early version of this manuscript.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Baker, J.E., 1982. Digestive proteinases of *Sitophilus* weevils (Coleoptera: Curculionidae) and their response to inhibitors from wheat and corn flour. Can. J. Zool. 60, 3206–3214.

Baker, J.E., Kramer, K.J., 1996. Biotechnological approaches for stored-product insect pest management. Postharvest News and Information 7, 11N–18N.

Barillas-Mury, C., Noriega, F.G., Wells, M.A., 1995. Early trypsin activity is part of the signal transduction system that activates transcription of the late trypsin gene in the midgut of the mosquito *Aedes aegypti*. Insect Biochem. Molec. Biol. 25, 241–246.

Bolter, C.J., Jongsma, M.A., 1995. Colorado potato beetles (*Leptinotasa decemlineata*) adapt to proteinase inhibitors induced in potato leaves by methyl jasmonate. J. Insect Physiol. 41, 1071–1078.

Broadway, R.M., 1995. Are insects resistant to plant proteinase inhibitors? J. Insect Physiol. 41, 107–116.

Bown, D.P., Wilkinson, H.S., Gatehouse, J.A., 1997. Differentially regulated inhibitor-sensitive and insensitive protease genes from phytophagous insect pest. *Helicoverpa armigera*, are members of complex multigene families. Insect Biochem. Molec. Biol. 27, 625–638.

Frugoni, J.A.C., 1957. Tampone universale di Britton e Robinson a forza ionica costante. Gaz. Chim. Ital. 87, 403–407.

Gatehouse, L.N., Shannon, A.L., Burgess, E.P.J., Christeller, J.T., 1997. Characterization of major midgut proteinase cDNAs from *Helicoverpa armigera* larvae and changes in gene expression in response to four proteinase inhibitors in the diet. Insect Biochem. Molec. Biol. 27, 929–944.

Gish, W., States, D.J., 1993. Identification of protein coding regions by database similarity search. Nat. Genet. 3, 266–272.

Hagstrum, D.W., Subramanyam, B., 1996. Integrated management of insects in stored products. Marcel Dekker, Inc, New York.

Hedstrom, L., Szilagyi, L., Rutter, W.J., 1992. Converting trypsin to chymotrypsin: the role of surface loops. Science 255, 1249–1253.

Houseman, J.G., Thie, N.M.R., 1993. Difference in digestive proteolysis in the stored maize beetles: *Sitophilus zeamais* (Coleoptera: Curculionidae) and *Prostephanus truncatus* (Coleoptera: Bostrichidae). J. Econ. Entomol. 86, 1049–1054.

Jany, K.D., Haug, H., 1983. Amino acid sequence of the chymotryptic protease II from the larvae of the hornet *Vespa crabro*. FEBS Lett. 158, 98–102.

Jongsma, M.A., Bakker, P.L., Peters, J., Bosch, D., Stiekema, W.J., 1995. Adaptation of *Spodoptera exigua* larvae to plant proteinase inhibitors by induction of gut proteinase activity insensitive to inhibition. Proc. Natl. Acad. Sci. 92, 8041–8045.

Jongsma, M.A., Peters, J., Stiekema, W.J., Bosch, D., 1996. Characterization and partial purification of gut proteinase of *Spodoptera exigua* Hübner (Lepidoptera: Noctuidae). Insect Biochem. Molec. Biol. 26, 185–193.

Johnston, K.A., Lee, M.J., Brough, C., Hilder, V.A., Gatehouse, A.M.R., Gatehouse, J.A., 1995. Protease activities in the larval midgut of *Heliothis virescens*: evidence for trypsin and chymotrypsin-like enzymes. Insect Biochem. Molec. Biol. 25, 375–383.

Kalhok, S.E., Tabak, L.M., Prosser, D.E., Brook, W., Downe, A.E., White, B.N., 1993. Isolation, sequencing and characterization of two cDNA clones coding for trypsin-like enzymes from the midgut of *Aedes aegypti*. Insect Mol. Biol. 2, 71–79.

Konarev, A.V., Fomicheva, Y.V., 1991. Cross analysis of the interaction of α-amylase and proteinase components of insects with protein inhibitors from wheat endosperm. Biokhimiya 56, 419–427.

Kraut, J., 1977. Serine proteases: structure and mechanism of catalysis. Ann. Rev. Biochem. 46, 331–358.

Langenfeld, J., Kiyokawa, H., Sekula, D., Boyle, J., Dmitrovsky, E., 1997. Posttranslational regulation of cyclin D1 by retinoic acid: a chemoprevention mechanism. Proc. Natl. Acad. Sci. USA. 94, 12070–12074.

Lee, M.M., Lee, S.H., Park, K.Y., 1997. Characterization and expression of two members of the S-adenosylmethionine decarboxylase gene family in carnation flower. Plant Molec. Biol. 34, 371–382.

Liang, C., Brookhart, G., Feng, G.H., Reeck, G.R., Kramer, K.J., 1991. Inhibition of digestive proteinases of stored grain coleopteran by oryzacystatin, a cysteine proteinase inhibitor from rice seed. FEBS 278, 139–142.

Maekawa, M., Sugano, K., Kashiwabara, H., Ushiama, M., Fujita, S., Ohkura, H., Kakizoe, T., 1998. Point mutations of ornithine decarboxylase gene are an infrequent event in colorectal cancer but a missense mutation was found in a replication error positive patient with hMSH2 germline mutation. Jpn. J. Clin. Oncol. 28, 383–387.

Marchal, C., Haguenauer, T.R., Urban, G.D., 1998. A PEST-like sequence mediates phosphorylation and efficient ubiquitination of yeast uracil permease. Molec. Cell Biol. 18, 314–321.

Matsumoto, I., Emori, Y., Abe, K., Arai, S., 1997. Characterization of a gene family encoding cysteine proteinases of *Sitophilus zeamais* (maize weevil), and analysis of the protein distribution in various tissues including alimentary tract and germ cells. J. Biochem. 121, 464–476.

Matsumoto, I., Abe, K., Arai, S., Emori, Y., 1998. Functional expression and enzymatic properties of two *Sitophilus zeamais* cysteine proteinases showing different autolytic processing profiles in vitro. J. Biochem. 123, 693–700.

Michaud, D., Cantin, L., Vrain, T.C., 1995. Carboxyl-terminal truncation of oryzacystatin II by oryzacystatin-insensitive insect digestive proteinases. Arch. Biochem. Biophys. 322, 469–474.

Müller, H.M., Crampton, J.M., Della Torre, A., Sinden, R., Crisanti, A., 1993. Members of a trypsin gene family in *Anopheles gambiae* are induced in the gut by blood meal. EMBO J. 12, 2891–2900.

Müller, H.M., Catteruccia, F., Vizioli, J., Torre, A.D., Crisanti, A., 1995. Constitutive and blood-induced trypsin genes in *Anopheles gambiae*. Exp. Parasit. 81, 371–385.

Murdock, L.L., Brookhart, G., Dunn, P.E., Foard, D.E., Kelley, S., Kitch, L., Shade, R.E., Shukle, R.H., Wolfson, J.L., 1987. Cysteine digestive proteinases in Coleoptera. Comp. Biochem. Physiol. 87B, 783–787.

Nielsen, H., Engelbrecht, J., Brunak, S., Von Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering 10, 1–6.

Oppert, B., Kramer, K.J., Johnson, D.E., MacIntosh, S.C., McGaughey, W.H., 1994. Altered protoxin activation by midgut enzymes from a *Bacillus thuringiensis* resistant strain of *Plodia interpunctella*. Biochem. Biophys. Res. Commun. 198, 940–947.

Paululat, A., 1996. Try29F, a new member of the *Drosophila* trypsin-like protease gene family, is specifically expressed in the posterior embryonic midgut. Gene 172, 245–247.

Peterson, A.M., Barillas-Mury, C.V., Wells, M.A., 1994. Sequence of three cDNAs encoding an alkaline midgut trypsin from *Manduca sexta*. Insect Biochem. Molec. Biol. 24, 463–471.

Peterson, A.M., Fernando, G.J.P., Wells, M.A., 1995. Purification, characterization and cDNA sequence of an alkaline chymotrypsin from the midgut of *Manduca sexta*. Insect Biochem. Molec. Biol. 25, 765–774.

Reeck, G., Oppert, B., Denton, M., Kanost, M., Baker, J., Kramer, K., 1998. Insect proteinases. In: Turk, V. (Eds.), Proteases. Birkhäuser, Basel, Switzerland. In press.

Sasaki, T., Hishida, T., Ichikawa, K., Asari, S.I., 1993. Amino acid sequence of alkaliphilic serine protease from silkworm *Bombyx mori*, larval digestive juice. FEBS Letters 320, 35–37.

Sinha, R.N., 1959. The hydrogen-ion concentration in the alimentary canal of beetles infesting stored grain and grain products. Ann. Entomol. Soc. Am. 52, 763–765.

Sreedharan, S.K., Verma, C., Caves, L.S.D., Brocklehurst, S.M., Gharbia, S.E., Shah, H.N., Brocklehurst, K., 1996. Demonstration that 1-*trans*-expoxysuccinyl-L-leucylamido-(4-guanidino)butane (E-64) is one of the most effective low Mr inhibitors of trypsin-catalysed hydrolysis. Characterization by kinetic analysis and by energy minimization and molecular dynamics simulation of the E-64-βtrypsin complex. Biochem. J. 316, 777–786.

Terra, W.R., 1988. Physiology and biochemistry of insect digestion: an evolutionary perspective. Brazil J. Med. Biol. Res. 21, 675–734.

Terra, W.R., Ferreira, C., 1994. Insect digestive enzymes: properties, compartmentalization and function. Comp. Biochem. Physiol. 109B, 1–61.

Titus, D., 1991. Promega Protocols and Application Guide. 2nd ed. Promega Corporation.

Wang, S., Hickey, D.A., 1996. Direct Submission to SWISS-PROT: locus TRYU_DROER, accession P54629.

Wang, S., Magoulas, C., Hickey, D.A., 1993. Isolation and characteriz-

ation of a full-length trypsin-encoding cDNA clone from the lepidopteran insect *Choristoneura fumiferana*. Gene 136, 375–376.

Wang, S., Young, F., Hickey, D.A., 1995. Genomic organization and expression of a trypsin gene from the spruce budworm *Choristoneura fumiferana*. Insect Biochem. Molec. Biol. 25, 899–908.